Towards an encoding model for Nandinagari conjuncts

Anshuman Pandey
pandey@umich.edu

April 25, 2017

1 Introduction

The Nandinagari script has been proposed for encoding in Unicode (see L2/16-310). The encoding model requires additional attention because of a method of conjunct representation that is not found in Indic scripts of the same general structure. This document describes the types of conjuncts used in Nandinagari and provides some models that may be used for representing these special conjuncts. The author requests that experts review the document and provide comments, such that the encoding for Nandinagari may progress.

2 Overview of conjunct forms

Nandinagari conjuncts are rendered in various ways. Usage of atomic and other distinctive ligatures is the most common method. As in Devanagari, some letters exhibit special behavior. The letter र RA is rendered as *repha* when C_1 and as *ra-kāra* when C_2. The letter य YA takes a post-base form when C_2, similar to Bengali. Half-forms of C_1 are used when the letterform allows for truncation of descenders.
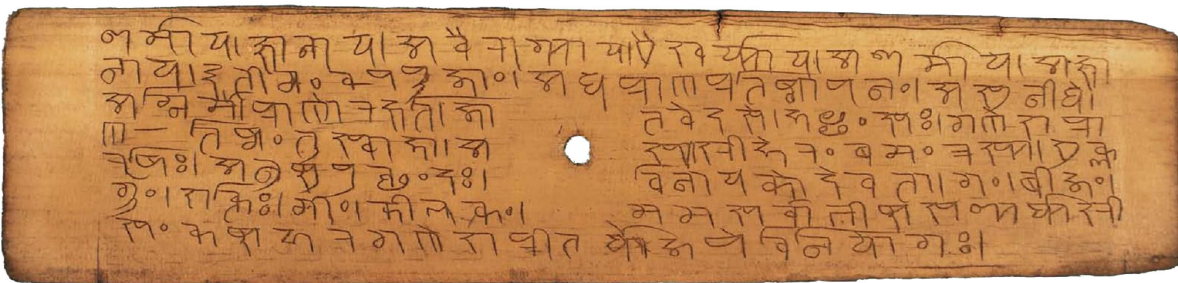
Lines 1 and 2 of the above folio read as follows:

dha·rmā·ya . jñā·nā·ya . a·vai·rā·gyā·ya . ai·śva·ryyā·ya . a·dha·rmā·ya . a·jñā·[-]
nā·ya . ? nī· ? pa·pū·jāṃ . a·tha·prā·ṇa·pra·ti·ṣṭhā·pa·nam . a·su·nī·the .

The following orthographic syllables found in the above lines illustrate the types of conjunct forms used in the script:

- ligatures: ज्ञा *jñā*, ष्ठा *ṣṭhā*

- half-form: श्व *śva*

- *repha*: र्मा *rmā*

- *ra-kāra*: प्र *pra*

- post-base *ya*: ग्या *gyā*

- *repha* and post-base *ya*: र्या *ryyā*

Another common method of representing conjuncts in Nandinagari is to join the consonants of a cluster using an extension of their headstrokes, or by causing each letter in the cluster to touch the right edge of the previous letter. In both cases, the constituent letters retain their nominal shapes. Such an orthography for conjuncts is meaningful because as a general convention, headstroke connections in Nandinagari are restricted to an *akṣara* (orthographic syllable) and do not extend to neighboring syllables. The headstroke connects vowel / consonant letters and spacing dependent vowels of an *akṣara*, while individual *akṣara*-s are separated by spaces (see the folio above). Following the conventions of the script, a group of letters connected by a single headstroke, or a group of letters with touching bodies is an indication of membership in an *akṣara* .
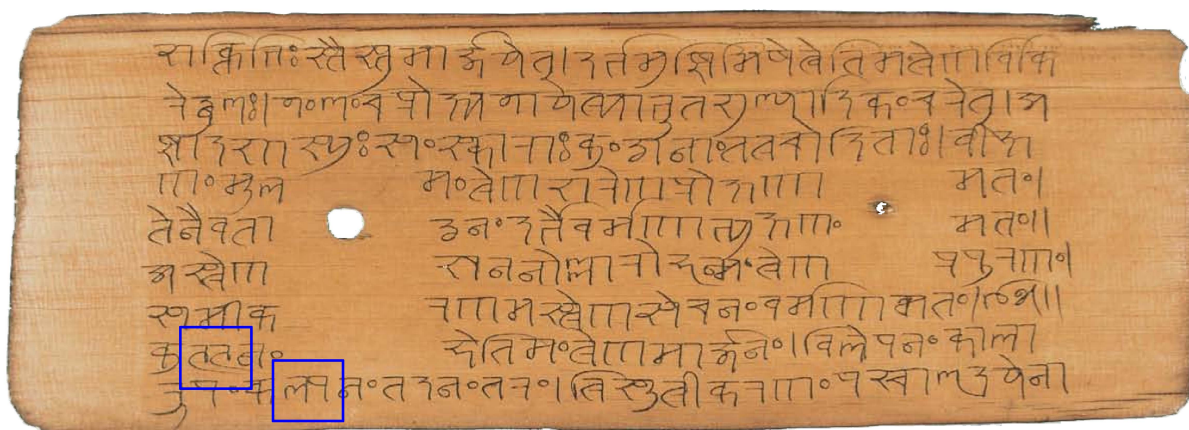
The above folio contains two touching conjuncts, shown in blue. The ट्ट *ṭṭa* occurs in क्ट्ट्नं *kaṭṭanaṃ* and the ल्प *lpa* occurs in क्ल्प्नं *kalpanaṃ*.

Most conjuncts can be expressed as either ligatures or as touching forms. For instance, the cluster द *da* + द *da* may be rendered as both द and द्द *dda*. In such cases, above-base and below-base non-spacing dependent vowel signs attach to the left-most letter. Top-right vowel signs attach to the top-right edge of the first letter. Left-side vowel signs naturally are placed after the left-most letter.

As these types of 'touching' conjuncts do not occur in Devanagari or scripts with headstrokes, an encoding model must be defined for Nandinagari, which enables representation of all forms of conjuncts.

3   Possible models

3.1   'Regular' conjuncts

The non-'touching' conjuncts, hereafter called 'regular' conjuncts, may be represented in Nandinagari by following the Devanagari model. These 'regular' conjuncts are rendered by default as atomic and distinctive ligatures, and in some cases using half-forms of C_1. These may be produced by placing VIRAMA after each letter in a cluster:

$$C, (\bigcirc\text{ VIRAMA}, C)^*$$

Examples of atomic and distinctive ligatures:

> *jña* ३ <ᘔ JA, ◌ VIRAMA, ᘒ NYA>
>
> *dda* ३ <ᘓ DA, ◌ VIRAMA, ᘓ DA>
>
> *lla* ᘲ <ᘪ LA, ◌ VIRAMA, ᘪ LA>
>
> *ṣṭa* ᘫ <ᘧ SSA, ◌ VIRAMA, ᘝ TTA>

A ligature may be formed using a half-form of C_1 if the letterform provides an opportunity for modification, ie. removal of the right descender. This type of ligature may be produced with the following letters: ग → ᘗ *ga*, त → ᘝ *ta*, म → ᘮ *ma*, शा → ᘢ *śa*, सा → ᘦ *sa*.

> *gna* ᘗᘠ <ग GA, ◌ VIRAMA, ᘠ NA>
>
> *śva* ᘢᘭ <शा SHA, ◌ VIRAMA, ᘭ VA>
>
> *ska* ᘦᘕ <सा SA, ◌ VIRAMA, ᘕ KA>

When य YA is final in a cluster it it represented by default using the post-base form ◌ᘯ:

> *gya* गᘯ <ग GA, ◌ VIRAMA, य YA>
>
> *tya* तᘯ <त TA, ◌ VIRAMA, य YA>
>
> *dmya* ᘓᘮᘯ <ᘓ DA, ◌ VIRAMA, ᘮ MA, ◌ VIRAMA, य YA>
>
> *yya* यᘯ <य YA, ◌ VIRAMA, य YA>

The letter ᘰ RA is rendered by default as ◌ᶜ *repha* when C_1, and as ◌ *ra-kāra* when non-initial:

> *rma* ᘮᶜ <ᘰ RA, ◌ VIRAMA, ᘮ MA>
>
> *pra* ᘩ <ᘩ PA, ◌ VIRAMA, ᘰ RA>

With the above rules for RA and YA, the following are the normal interactions for the letters when they are occur together in a cluster:

3

rya	र्य	<र RA, ्् VIRAMA, य YA>
rṣya	र्ष्य	<र RA, ्् VIRAMA, ष SSA, ्् VIRAMA, य YA>
prya	प्र्य	<प PA, ्् VIRAMA, र RA, ्् VIRAMA, य YA>

3.2 'Touching' conjuncts

There are a few feasible encoding models for 'touching' conjuncts:

1. *Non-final consonants as conjoined forms* There does not appear to be a requirement to render special conjoining forms of the initial letter in a cluster. Therefore, it may be possible to consider the initial and all non-final letters as 'conjoining' forms in a 'touching' conjunct with the final letter as the 'base'. In Devanagari, the conjoining form of C_1 is requested by placing [ZWJ] U+200D ZERO WIDTH JOINER (abbreviated as ZWJ) after VIRAMA. The meaning of the sequence would change in Nandinagari:

$$C, (\text{ VIRAMA}, \text{ ZWJ}, C)*$$

   It would be used as follows:

dda	द्द	<द DA, ्् VIRAMA, द DA>
dda	दद	<द DA, ्् VIRAMA, [ZWJ] ZWJ, द DA>

2. *Non-initial letters as conjoined forms* Treat each non-initial letter in a cluster as a conjoining form. Following the specification for Oriya, it may be possible to use [ZWJ] U+200D ZERO WIDTH JOINER (abbreviated as ZWJ) before VIRAMA for producing conjoined forms:

$$C, (\text{ ZWJ}, \text{ VIRAMA}, C)*$$

   Accordingly:

dda	द्द	<द DA, ्् VIRAMA, द DA>
dda	दद	<द DA, [ZWJ] ZWJ, ्् VIRAMA, द DA>

3. *New control characters* Encode a special VIRAMA-like character for specifically producing 'touching' consonants. For sake of discussion, this character is given the generic shape [◌̣] and the name 'VIRAMA-2'. It would be used as follows:

dda	द्द	<द DA, ्् VIRAMA, द DA>
dda	दद	<द DA, [◌̣] VIRAMA-2, द DA>